DOCUMENT RESUME

ED 446 050 SP 039 481

AUTHOR Smagorinsky, Peter
TITLE What English Educators Have To Say to Assessment Specialists.
PUB DATE 2000-00-00
NOTE 11p.
PUB TYPE Opinion Papers (120)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *English Instruction; English Teachers; Evaluation Methods; Higher Education; Language Arts; Language Usage; Literature; Standardized Tests; *Student Evaluation; Writing Skills

ABSTRACT

This paper discusses the hazards of using standardized assessment in English/language arts. Using standardized tests to measure student competency with language is problematic because of inattention to what counts as appropriate language usage. Standardized tests reify textbook language usage only and do not distinguish between correct and stylistically appropriate language. Assessment of literature typically involves a brief passage from a story followed by questions asking the student to identify themes or key facts. Though it is important to be able to make sense of a text before doing other things with it, it is those other things that are primarily beneficial in reading the literature. Those other things would be difficult to assess with conventional standardized testing. Writing is not always part of the assessment because it cannot be machine graded. Grading is done by a cadre of raters who score papers in reasonably similar ways. However, good writing in one context might be viewed as bad in another. It is questionable to claim that a student's production in response to a prompt will yield writing that raters can sensitively evaluate according to a common rubric. Also, quality writing usually occurs over time, not in a high-pressure situation. (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

What English Educators Have to Say to Assessment Specialists

Peter Smagorinsky

University of Georgia

125 Aderhold Hall

Athens, GA 30602

smago@peachnet.campuscwix.net

ED 446 050

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

P. Smagorinsky

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

SP039481

BEST COPY AVAILABLE

I should begin my perspective with a few qualifiers. First of all, I don't presume to speak for my field. My field, broadly known as English/Language Arts, is divided on many of the issues I'll address. While my views may resonate with those of many, they may well seem preposterous to others. There is an ample empirical basis to support the views I'll present, though people who disagree with me could easily cite an impressive array of studies to the contrary. And so my remarks should be understood as one perspective amidst many, though I hope one that raises questions for my colleagues in the field of measurement.

The second area I should delimit is the nature of my field, English/Language Arts. Traditionally this field has included three major curricular strands: language, literature, and writing. And indeed that is how I will organize my remarks, since these categories often serve as arenas for measurement on standardized assessments. The categories, however, are growing obsolete as both technology and literacy research suggest the need for the inclusion of multiple literacies--including literacy in computers, the arts, popular culture, and other modern tools—in the realm of English/Language Arts. Additionally, attention to interdisciplinary knowledge suggests that relationships among subject areas are important in becoming a thoughtful, literate citizen: the historical and social issues that motivate literature, the interrelatedness of arts in characterizing the human condition, relationships between the arts and science such as literature that takes an environmentalist position, and so on. My effort to delimit my field thus results in quite the opposite, the position that my field potentially has no limits. I imagine that this perspective could cause some difficulties for anyone charged with the task of measuring it.

For the purposes of convenience, I'll sort my comments into the three traditional strands of my discipline. In discussing them, however, I'll try to demonstrate the limitations of thinking about my field in strict categories.

Language

My familiarity with the measurement of competence with language comes from my many years of taking standardized tests and my many years of preparing students to take them. Typically, students are asked to do multiple choice items in the areas of analogy, vocabulary, and usage.

My trouble with this approach to measuring competence in language is the inattention to the question of what counts as appropriate usage. Standardized tests reify textbook language use, what's often called standard English. However, for many years now linguists, anthropologists, communication researchers, educational discourse analysts, rhetoricians, and others have extensively documented the ways in which the notion of correct vocabulary, syntax, and usage are situational. From this perspective, a standard form of a language is the form that accomplishes local rhetorical purposes and needs. In school, textbook grammar is authoritative, at least on tests in English classes. It is rarely actually spoken, however, even by faculty. It is certainly not the norm in memos issues by school administrators, try as they might.

I'll give a few examples of what I mean. Musicians performing within particular genres—country and western, rhythm and blues, and rock and roll—use a syntax that, according to textbooks and standardized tests, is incorrect. However, it is the appropriate syntax for use within the genre; the use of textbook grammar in these contexts for some lyrics would have a detrimental effect on the artist's success. A person with linguistic

competence, from this perspective, is not one who always uses textbook grammar but one who has a repertoire of discourse genres and social languages at his or her disposal and knows how and when to use each. Such a person is savvy about what's called code-switching; that is, having the communicative competence to know when particular language codes should be invoked. At an interview with IBM, textbook grammar is called for, while closing an oil deal in Oklahoma might require proper knowledge of when to say "Y'all ain't fixin to do nothin funny now."

English teachers tend to focus on teaching textbook grammar, usually in the highly limited context of doing grammar lessons. In my view, any instruction and assessment in language use ought to focus primarily on how to use language in context. This effort would require the recognition that the notion of a language standard is local, rather than being applicable to textbook grammar only. It would also recognize that every version of a language is a dialect, not just those that depart from textbook grammar. Textbook grammar, then, is a dialect that is appropriate and effective in some situations but probably would cause problems in others. Terms such as African American Vernacular English become questionable because of the implication that as a vernacular it is a deviant form of proper English, when in some situations it is proper English—just ask Eddie Murphy and other entertainers who've gotten rich using it.

In terms of using language, it's also important to distinguish between what's correct and what's stylistically appropriate. Much correct language use—that is, following the rules of textbook grammar—is stylistically inappropriate, such as the use of excessive passive voice constructions in a narrative designed to be lively. Furthermore, notions of correct style vary by genre, something overlooked by those who design

grammar and usage computer programs. The speeches of Dr. Martin Luther King, Jr., which are thought to be the greatest instantiations of the African American oratory style, are red-flagged for redundancy by these programs. Again, given the situational nature of appropriate language use at the sentence- and genre-levels, measuring students' facility with a single, rarely-used dialect--textbook grammar—reflects a misunderstanding of what laypeople and scholars have known about language for some time.

Taking this perspective would require substantial changes in how language competence is measured in standardized tests. Rather than measuring knowledge of textbook English—which favors those in whose homes something proximate to textbook English is spoken—students' knowledge of multiple language codes and their appropriateness in particular social settings would be assessed. And so any language assessment item would need to be prefaced by a context. I imagine that it would be difficult to come up with a way, however, to state unequivocally that one answer is right and all others are wrong. But that's a conundrum for my colleagues in measurement.

Literature

Assessment of literature typically comes in the form of a brief passage from a story, followed by a set of questions asking the student to identify the main theme or most appropriate title, some key facts about the passage, and so on. The assumption behind this approach is that literary understanding involves the ability to state correctly what has happened in the text. Undoubtedly, the ability to make sense of a text on the literal level is prerequisite to doing other things with it. However, those other things are, I would argue, what is primarily beneficial in the reading of literature. Unfortunately, I imagine that they would be very difficult to assess on a conventional standardized test.

First I'll discuss what I call the codification of texts. Texts are codified on at least two levels. One is the word level, which seems to be the emphasis of most measurements of literature I know of on standardized tests. That is, the role of the student is to decode the words and know how they function to relate a narrative. Again, if you can't do this, then you probably can't do much else, so I don't wish to say that assessment at this level is entirely inappropriate.

Texts are also codified at the level of genre. Knowledge of genre-level codes enables reading at a more inferential level and suggests how words, particularly as they work in conjunction with one another, should be interpreted. For example, a text can be coded in a variety of genres: an argument, a satire, a Western, and so on. At times a text can include multiple codes. I'd say, for instance, that Thomas Berger's novel Little Big Man is a satirical Western that argues against American notions of progress. Knowledge of textual codes and conventions can be important in understanding what an author is trying to say, if that's what you're trying to do. The late Chicago columnist Mike Royko, for example, was substantially on record as being a proponent of gun control, in part because he'd been held up at gunpoint by a young thug who could easily have killed him, an incident he'd written about in his column. On a later occasion, he wrote a column in which he stated that the NRA was too wimpy in its stand that assault rifles should be legal; Royko argued that citizens should be able to carry around bazookas and armed nuclear missiles. A number of people subsequently wrote letters to him saying that he was a hypocrite for supporting gun control on the one hand and advocating heavy citizen armament on the other. I would say that they missed out on the ironic codes of his column, a view that Royko himself supported in his own comments on the incident.

A second important consideration in assessing literary understanding is, I think, entirely overlooked in standardized assessment, that being the reader's constructive activity during the process of reading. While much school-based instruction and standardized assessment focuses on the presumed meaning of the text itself, a great many reading researchers focus on what readers do with texts while reading. It's now well-documented that different readers bring different experiences to texts that make the meaning that they construct quite different, even while being responsive to the same set of textual codes. In my own work I have found that these idiosyncratic responses are enhanced when students are allowed to choose their own medium of interpretation. I've studied students using art, dance, drama, and different forms of writing to interpret literature, each with its own potential for enabling students to represent their understanding of the text. I imagine that it would be difficult to build artistic interpretation into standardized assessment. Nonetheless, I've found it to be a medium with great potential for enabling diverse students to produce compelling interpretations of literature.

Furthermore, the context of reading may suggest different kinds of responses. At a book club consisting of friends who read and meet for enjoyment, laughter and tears are legitimate ways of responding to a text. In a school classroom or standardized testing situation, however, crying in response to a text would likely be considered inappropriate; it would surely gain you no points on the SAT. I would say, however, that assuming that the testing context is either neutral or irrelevant to measurement overlooks the abundant work in cultural psychology that documents the ways in which cultural practices and contexts channel cognition. Assuming the neutrality or irrelevancy of the testing context

undoubtedly makes the assessment more congenial to those who find it familiar and comfortable, and penalizes those who find it alien. I would say that giving some attention to the effects of context on reading would lend authenticity to the scores that result from the assessment.

Writing

Writing is not always part of the assessment mix because it can't be machine-graded. Writing samples need to be scored by a cadre of graders who must be trained according to a rubric so that reliability is assured across writers and raters. That's an expensive proposition, both in terms of training and scoring. In many cases, then, this rather critical strand of English/Language Arts is eliminated from consideration.

I manage a family budget and know that you can't have everything you want, so I understand the financial problems involved in assessing something that's expensive to administer. For now, then, I'll focus on problems associated with the kinds of large-scale writing assessment that I'm familiar with, those that train a large number of raters to score papers in reasonably similar ways in order to assure reliability.

Again, I'll start by reviewing some of the issues in my field, or at least issues that I think are important. One is the recognition that there are disciplinary conventions for writing. In other words, all writing does not follow the same conventions; rather, writers in different fields, and different communities of practice within the same field, adhere to different conventions in producing effective texts. Writing in the hard sciences, for instances, tends to favor precision and specificity, with a technical vocabulary that most readers would not characterize as belletristic. Agency is attributed to the phenomena under study, rather than to the scientist who studies it, leading to frequent passive voice

sentence structures. Postmodern scholars, in contrast, emphasize the subjectivity of observation and locate agency in the observer as well as the observed. These two different disciplinary perspectives result in very different values in how one writes, even when the same basic genre, such as an argument, is being constructed by both.

The point here is similar to the point I tried to make when discussing the assessment of language use: that writing is a communicative act and that effective communication involves knowledge of which conventions to invoke. Good writing in one context (a legal brief) might be viewed as bad writing in another (a love letter). It's highly questionable, then, to claim that a student's production in response to a prompt will yield writing that a cadre of raters can sensitively evaluated according to a common rubric. The most important aspect of the writing—whom the student is intending the writing to reach—is not a factor in assessments of this type. Again, communicative competence is the central issue, as manifested in knowledge of discourse conventions. What matters for writers is how broad their repertoire of conventions is and how astute they are in knowing when to apply each set.

A second problem of assessing writing in a single-session, high-stakes measurement is the well-established idea that writing usually takes place over an extended period of time, rather than in a pressure-cooker setting. While some writers do indeed write under great pressures, such as news reporters writing against daily deadlines, much writing is accomplished as a longer process that might include jotting notes, going for a walk, writing a section, folding laundry, writing a bit more, pausing to read, revising the first draft, taking a nap, etc. This might take hours, days, weeks, or longer. Compressing writing into a finite period, on a topic of perhaps limited interest to writers,

without a specified readership, in a setting that favors some writers over others, and using a rubric that applies equally to all regardless of differences in students' construction of the situation, is, I would argue, a questionable way of measuring the absolute quality of student writing.

Conclusion

Like most, I am a better critic of measurement efforts than an architect of alternatives. I recognize that standardized tests are inevitable in American culture, while also noticing that other countries do just fine without them. I do hope, however, that the points I have tried to make about my field have helped my colleagues in measurement understand at least one person's perspective on what it means to be literate and consider these issues in their work.